

Swedish Employment in the 1950s – How to Fill the Lacuna

Gudmundur Gunnarsson¹ and Thomas Lindh²

We present a case study of how to link historical time series. There are no published data on employed people in Sweden between 1950 and 1960. To fill this lacuna we used unpublished extrapolations based on hours worked and provide consistent links for major industries. As an alternative linking series we use taxation statistics on active income earners adjusted to bridge a substantial change in method and definition between 1950 and 1951. Our work illustrates several common problems concerning reclassifications, definition changes, documentation deficiencies and changes in data collection. In validating the data we have used a wide range of information: cross-checking extrapolations, recompilations of censuses, correcting errors, comparisons between sources, and contemporary business cycle information. Our work shows that employment increased in the 1950s, with close to even numbers in the private and public sector, contrary to the strong concentration on public employment growth that previous data indicate.

JEL codes: C80, N34

Key words: Industry classification links; taxation statistics; historical time series.

1. Introduction

The linking of historical time series is a precarious business. Source data are often scarce, of dubious quality and not always well documented. Changes in definitions, collection methods and classifications compromise the comparability of data. We present a project to bridge the gap between historical employment series 1870–1950 and the Swedish National Accounts estimates 1960 and forward. Our reconstruction provides a case study illuminating common problems like poor documentation, differences in industry classifications and definitions between sources and over time, differences in data collection, and the choice of appropriate benchmarks.

From a methodological viewpoint our work does not rely on any single principle. Instead we take advantage of different information sources in a pragmatic way and use comparisons to validate our results. We also acknowledge that there is no ‘best’ standard

¹ Department of Business Studies and Informatics, Mälardalens University, Box 883, SE-721 23 Västerås, Sweden.

² Department of Economics, Uppsala University, Box 513, SE-751 20 Uppsala, Sweden and Institute for Future Studies, Box 591, SE-101 31 Stockholm, Sweden. Correspondence: Thomas Lindh. Fax +46-18-471 1478, Phone +46-18-471 1103. E-mail: thomas.lindh@nek.uu.se

Acknowledgments: We are grateful to Michael Wolf at Statistics Sweden for his help with the source material for the extrapolated employment series and to Ragnar Bentzel for the initial purchase of these data and permission to use them. We thank Pebbe Selander both for his valuable comments on the text and for providing us with unpublished research material on the taxation statistics. We are also grateful to Olle Krantz and referees for helpful comments and advice.

series for all purposes. Different alternatives may be preferred depending, for example, on whether the research questions concern variations or levels of employment, or whether the industry classifications are crucial or not.

Therefore we do not present any final preferred estimate of employment in the 1950s. The information we provide must be used with discretion in order to avoid misleading conclusions. Researchers must make an informed choice adapted to the purpose of each study. Still, although the two main alternatives we provide for filling the lacuna differ substantially, both of them are better than any published alternatives. We can conclude with reasonable certainty that employment grew over the 1950s by around 250,000 people. This is much more modest than straight comparisons between historical and modern series would lead us to believe but it is not as small as the difference between the 1950 and 1960 Censuses indicates nor was employment growth as concentrated in the public sector as straight comparisons between the censuses imply.

Below is an overview of our sources and methods. Section 2 discusses methodological issues in detail. In Section 3 the validity of our estimates is discussed. Concluding remarks are in Section 4.

1.1. Background and overview

No comprehensive statistics on Swedish employment in terms of individuals were collected in the 1950s. Statistics Sweden (SCB) reconstructed data on hours worked 1950–1963. Employment figures from Labour Force Surveys (AKU) start in 1960. Jungenfelt (1959, 1966) constructed annual employment series for major industries 1870–1950. Taxation statistics 1943–1966 are also available.

In addition to these sources we use the 1950 and 1960 Censuses and the National Accounts. These three sources are used for calibrating and benchmarking. We use contemporary employment estimates from the National Institute of Economic Research (KI) for validation. The diagram in Figure 1 summarizes our sources and corrections.

All of the above data sources had different purposes when collected and constructed. They also pose different problems in scope and methods. Let us therefore specify each data source:

- a) Jungenfelt's purpose was to study economic growth and structural change 1870–1950. He extended older series using a hybrid of the classification in the taxation statistics and the Swedish SNI 1960 code. His data from the 1940s are to a large extent from the taxation statistics, but complemented by other sources and benchmarked to censuses.
- b) Taxation statistics on active income earners (SOS) were collected to produce income and wealth statistics. Individuals are classified by nine (eight before 1952) major industries and whether self-employed, employed or not active. Although this is no standard employment definition,³ active income earners reflect changes in employment reasonably well.
- c) The published SCB series on hours worked was derived from income statistics

³ Temporary and seasonal employment is included to a greater extent in the case of active income earners. The industry definitions are closely related to those used by Jungenfelt and the 1950 and 1960 Censuses. However, they are based on self-assessments.

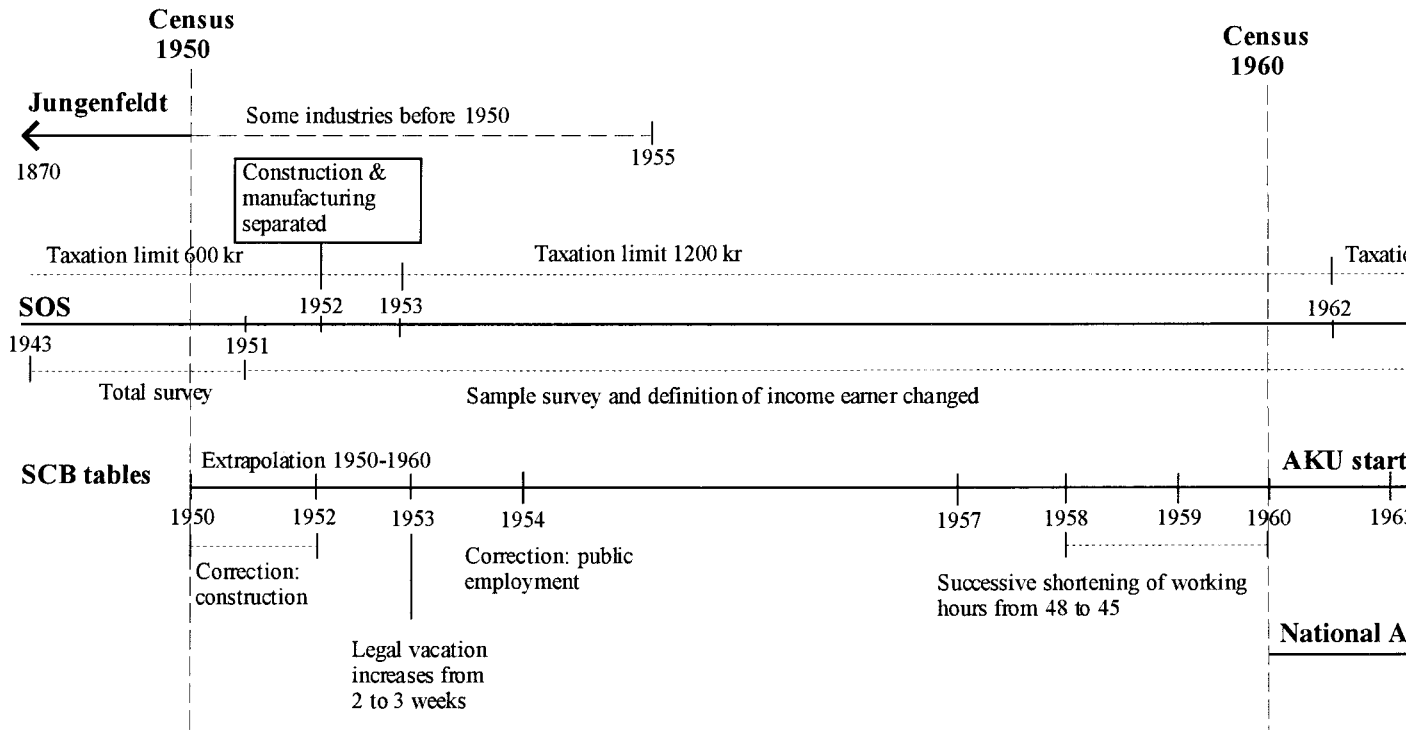


Fig. 1. Overview of the main data sources, problems and corrections

except for manufacturing and agriculture where industry surveys and statistics were available. The old Swedish industry classification (SNI) was replaced in 1960 with SNI 1960 that is closer to the current ISIC standard. The 1950–1963 code differs from this mainly with respect to some public employment industries. The hours series was the basis for later unpublished extrapolations in 1985 to employed individuals which have circulated among researchers.

- d) The 1950 and 1960 Censuses report the number of gainfully employed. Gainfully employed refers to a point estimate.⁴ Most seasonal work is thereby excluded. The number of the gainfully employed, therefore, understates the number of active income earners, as well as employed individuals according to AKU.
- e) KI published estimates and/or forecasts of employment changes at somewhat irregular intervals, starting in the 1940s. The industry classification is similar to that of the SOS series.

For research purposes a linking to the historical time series should satisfy three criteria:

- Levels of employment in 1950 and 1960 should be comparable as far as possible.
- Variation in employment should not be misleading compared with other historical series.
- The classification with respect to employment status and industries should be comparable.

The primary task is to get the levels approximately right, otherwise it automatically distorts the other two criteria. For the relation between employment and other variables the second criterion is important. Errors in the timing of changes can completely change the results of a regression analysis. Any analysis of industry employment and the structural change, of course, relies on the third criterion in order to avoid faulty conclusions.

But the criteria are either more or less important relative to each other depending on the exact research questions to be answered. For example, a regression analysis of short-run productivity changes and causal connections in the whole economy would depend crucially on the second criterion being met, while the first and third would be less important. An analysis of long-run productivity growth, on the other hand, would mainly depend on the first criterion.

It is, moreover, difficult to satisfy these criteria to the same degree at both ends of the interval. For instance classification is simpler to reconcile between different SCB series than it is to reconcile between Jungenfelt and the SCB series, while the SOS series is simpler to link to the Jungenfelt classification. Our choice was to focus on the level criterion and only provide the necessary information for researchers to handle the variability in a way suited to their purposes.

The two main sources for bridging the employment lacuna are the extrapolation from hours worked and taxation statistics. We reclassify industries and correct errors in the original SCB series. The SOS series pose two main problems: changes in survey

⁴ To be gainfully employed also requires substantially more work (half the regular work-time) than the AKU definition (one hour a week). In the 1950 Census only work on 1 November qualifies, in the 1960 Census work in the week before is sufficient.

techniques and definitions, and changes in taxation limits. The change in survey techniques and definitions cause a jump in the data 1950 to 1951 that we attack in two steps. First we correct the total between surveys and then we deal with the major problem: distributing the difference over industries and occupational status. We use the 1950 and 1960 Censuses to reclassify industries to the modern SNI 69/77. Finally we check the variations of the SOS and the SCB tables by means of business cycle data from KI.

Our two alternatives have different advantages and drawbacks, depending on whether you are mainly interested in the older part of the series or the post-war part, and the level of classification you want. They also have different problems with regard to variation: the SCB series is unreliable concerning the timing of changes while the SOS series has two artificial jumps and a bias in the growth rate. The levels we find after correction are, however, not too different. In respect to that criterion, they are therefore more or less equivalent. In the following we try to keep the discussion general to avoid burdening the reader with unnecessary details. Gunnarsson and Lindh (1997) offer a more detailed account of our work, including most of the data used.

2. Linking Strategies and Problems

The first problem was how much employment had increased over the 1950s. The difference between Jungenfelt's 1950 estimate of 2,977,500 employed and the 1960 employment estimate in the National Accounts (3,615,700) indicates an increase by more than 20 per cent. The increase in the number of the gainfully employed between the 1950 and 1960 Censuses – from 3,104,756 to 3,244,084 – indicates a modest growth of less than five per cent. The second problem was to bridge differences in classification of industries and employment status. We treat these problems separately to avoid confusion.

Below we first give details of how the SCB and the SOS series were constructed. After that we focus on how levels were inferred. Finally the derivation of classification links is discussed.

2.1. Understanding the data

The first task was to understand the construction of the data. The published taxation statistics had detailed source notes that were very helpful but still could not answer all questions. The unpublished extrapolation was documented by hand-written comments that were harder to decipher.

From 1943 and onward annual income data were collected by SCB from tax assessment records to produce income and wealth statistics.⁵ This is the main source of raw data for both of our alternative employment series. While the SCB series mainly are based on the tax-assessed income data the SOS series are based on the income earner data.

⁵ Up to 1966 these data were reported in the publication series Official Statistics of Sweden, Income and Property, Tax Assessments (SOS Inkomst och Förmögenhet, Skattetaxeringen). For simplicity we refer to these publications as SOS even though they also originate from SCB. Note that for example SOS 1952 contain data that refer to income earned in 1951 and so on. In 1967 comprehensive changes in method and definition were introduced and the data were thereafter reported in other publications.

Table 1. Index for the legislative change in annual hours compared to 1960

Year	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
Index	108.8	108.8	108.8	106.7	106.7	106.7	106.7	106.7	104.4	102.2	100

2.1.1. The SCB series

The industry classification in the SCB series is by an older SNI code that was replaced in 1960.⁶ At the one-digit level the SNI 1960 code is identical to the SNI 69 code.⁷ We only aim to estimate employment at the one-digit level, so this is no problem. The classification in the 1950–1963 series differs mainly with respect to the classification of public employment industries. This is problematic since we want to separate public and private employment to conform to the SNI 69/77 convention.

The published series 1950–1963 on hours worked are mainly computed from wage sum statistics in the following way (Statistical Reports N 1975:98, pp. 13–15). To get the hours worked by employees per year, the total wage sum in each industry was divided by the estimated average hourly wage rate in that industry. These data were checked against and supplemented by some independent sources on hours worked. Very uncertain estimates of self-employed hours have been added. The published series are not reported by occupational status as an explicit acknowledgement of this weakness.

In 1985 an extrapolation from these data to employed individuals was undertaken at SCB. The method used was to divide industry hours per year by stylised annual hours of work for a hypothetical person in each industry. The first step in computing stylised annual hours of work was to establish an index of legislative decisions on shorter working time (see Table 1).⁸ The index was then used to extrapolate backwards the 1960 level of average annual hours of work in each industry.

The stylised annual hours of work are thus based on two assumptions. First, the actual changes in the regular work time in each industry – including self-employed hours – follow the changes in the law. Second, that the relative spread of hours per year and person across industries, occupation and in the public and private sector is the same during the whole decade.

None of these assumptions are innocent a priori. The index construction assumes legislation to have a discrete effect on annual hours. This is liable to overstate the number of employed at the time of legislative change, since actually worked annual hours probably adapt more smoothly, both before and after the change. The industry spread in annual hours of work in 1960 is quite considerable. The reported number of hours worked per year and individual is e.g., 1,200 in agriculture compared to 2,900 in private services. Agriculture has a high proportion of seasonal work yielding low annual hours, but also a high share of self-employment where hours are estimated by surveys.

2.1.2. The SOS series

The taxation statistics contain a major stumbling block for both level and classification links. In 1951 a radical change in definition coincides with data collection being changed

⁶ A conversion table between these classifications is available in the same publication as the estimates: Statistical Reports (Statistiska Meddelanden) N 1975:98, Appendix 5, Table 98, pp. 59–60.

⁷ The SNI 69 code is defined in Reports on Statistical Co-ordination 1969:8. See also Table 2 below.

⁸ Silenstam (1970) details some further information about the work-time shortening in the 1950s.

from a total to a sample survey with no overlap. This change was implemented the year after the Census in 1950 that provided the benchmark for the Jungenfelt data. The SOS total of 3,007,217 active income earners in 1950 is pretty close to Jungenfelt's employment figure. Gainful employment in the census underestimates the number of active income earners by around ten per cent.

The basic problem with the SOS series is that in 1943–1950 the data refer to the total population of tax records. 1951–1966 the data were collected by sampling about a tenth of the total population but retaining a total survey of the highest income brackets. This is a considerably more comprehensive survey than AKU uses so the sampling by itself causes little concern. Two other features complicate the matter, however. First, there is uncertainty concerning the coverage of the sample and the appropriate multiplier to use for population totals. Second, and most serious, the shift was associated with major changes in definition: in particular co-taxed couples were counted as one individual in the total survey but as two separate individuals in the sample survey.

Another problem is the lower income limit for taxation. This limit was 600 SEK up to 1952, 1,200 SEK 1953–1961, so anyone earning less is excluded. This has two consequences. First, in the years 1953 and 1962 when the income limit is raised a considerable number of income earners in the lowest income bracket are suddenly excluded. Thus there are artificially induced slumps in the data. Second, the constancy of the nominal limit exaggerates the rate of growth of active income earners over the period of constancy, since a number of very low-income earners are added by inflation.

There are some other minor differences before and after the change in method. First, in the total survey before 1951 undistributed estates of deceased individuals and family foundations were counted as nonactive income earners. From 1951 only individuals who were alive at least part of the year are included. This affects only nonactive income earners and is therefore disregarded.

The source texts warn that the coverage of the post-1950 total survey of taxable incomes above 30,000 SEK is suspected of being incomplete. The locally elected chairmen of the taxation boards collected these data during a rather short and intense period (when final tax assessment may not have been available).⁹ But there are only around 15,000 active income earners above the 30,000 SEK limit so this is negligible in the context of employment estimates.

2.1.3. Summary

The problems in the two sources are very different. Determining the level change over the 1950s for the SOS series is mainly a question of bridging the 1950–1951 change in definition and data collection, while the level in the SCB series needed correction of calculation errors and validation against other information from the period. The SOS series classify individuals into an old industry definition close to that used by Jungenfelt, while the SCB series classifies industries more or less as modern National Accounts, with the exception of the distinction between public and private employment. Our main classification

⁹ Pebbe Selander and Roland Spånt – in unpublished research material about income distribution – estimate this loss of coverage at 15–30 per cent in connection with a change in the income limit from 30,000 SEK to 50,000 SEK in 1961. Our comparisons before and after 1951 indicate that the loss was considerably less as sampling started.

problem is therefore to link the SOS series to the National Accounts in 1960, and the SCB series to the Jungenfelt definitions in 1950.

2.2. *Getting the levels right*

A look at the data is valuable in order to get an impression of which problems we need to deal with and which can be ignored. In Figure 2 the total number of active income earners 1943–1966 is compared with National Accounts employment 1960–1969¹⁰ and Jungenfelt's employment 1940–1950. The positive growth bias is noticeable, especially in the 1960s, but not very prominent. The change in income limit 1961–1962 is clearly discernible as a downward jump without counterpart in the National Accounts, while the corresponding 1952–1953 downward jump is hard to discern.

The prominent jump in SOS 1951 as the total survey becomes a sample survey is what catches the eye, however. For purposes of level links we can live with the income limit bias, if we link in 1960. The jump 1950 to 1951 must, however, be dealt with in order to link to historical series. In Figure 2 is also the total of the (corrected) SCB series. The SOS series and the SCB series move similarly but with a lead for the extrapolated data that is likely to be an artefact.

There are around 100,000 self-employed income earners less in the SOS series than in the National Accounts. The discrepancy slowly decreases with time. This is explained by the National Accounts definition, which includes family members working with or without pay in family firms. By SOS definitions family helpers are either employees or not reported. However, it is only in agriculture that family helpers are substantial numbers.

On the other hand, there are more active income-earner employees than employees as defined in the National Accounts. The bias from lower tax limits is highly concentrated in this part of the series, reflecting that SOS employees to a larger extent include low-income earners with irregular or seasonal work than the National Accounts employees do.

2.2.1. *Recomputation of the SCB series*

The SCB series on hours worked include public services, which we wish to separate from private employment. Hours worked in public employment, defined as those worked by employees of state and local authorities in 1960, were therefore subtracted from the hours series. We assume that the 1960 proportions of public employment hours are the same over the whole decade. Public employment is highly concentrated in SNI 9, so except for this industry no large errors are generated.

After subtraction of public employment annual hours of work in each industry were computed. To check the extrapolations we divided hours worked in each category by the extrapolated employed individuals according to SCB to get back to annual hours of work. Errors were spotted by checking this against the extrapolation of annual hours of work. We found one very large discrepancy in the construction industry, SNI 5, 1950–1952 (around 100,000 people too many), another in public employment 1954 (around 10,000 people too few), and some minor ones. The first discrepancy was quite important for the relative development of public and private employment (see Figure 3).

¹⁰ SM N 1975:98 Appendix 5, table 43.

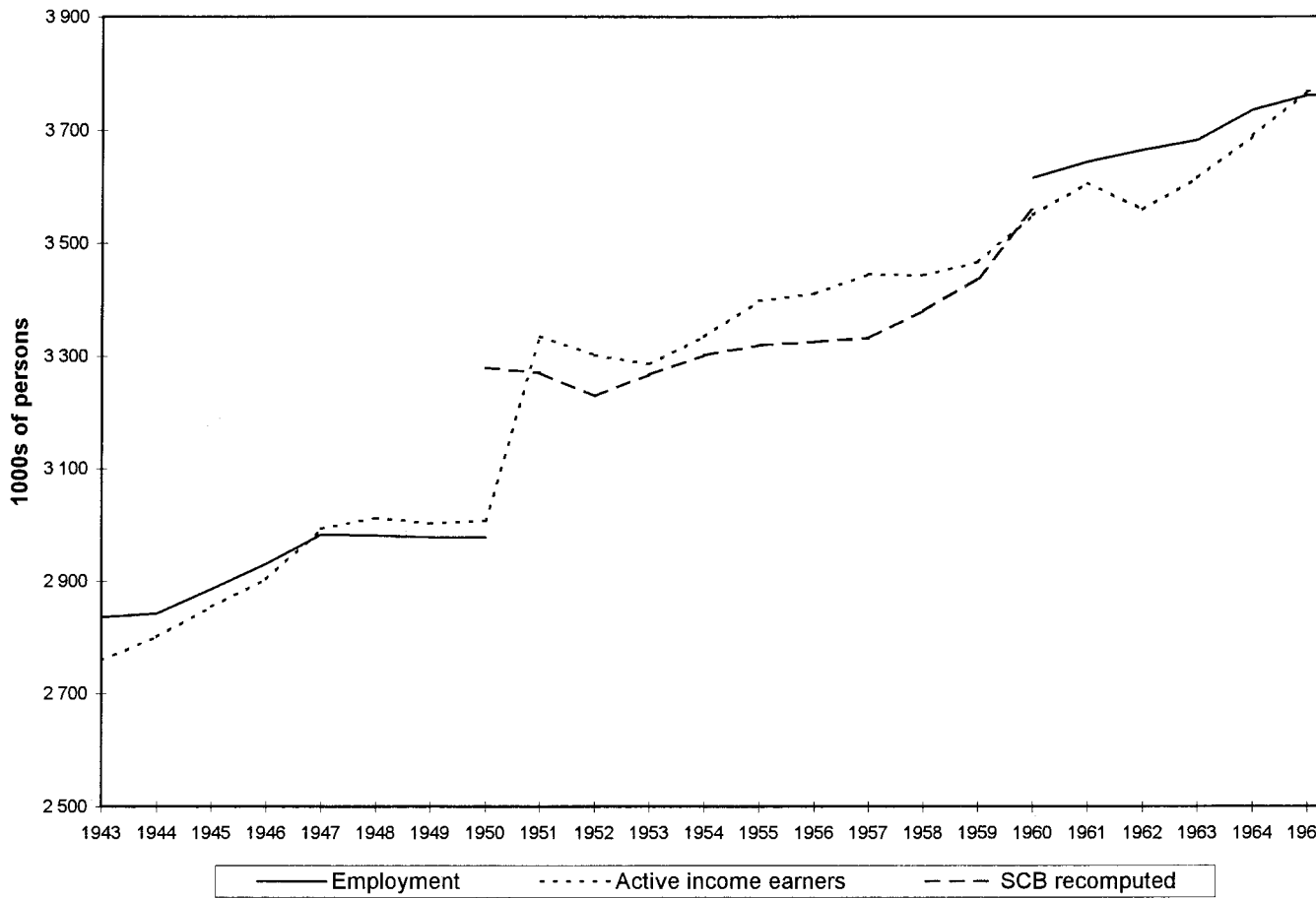


Fig. 2. Employment 1960–1969 National Accounts, 1943–1950 Jungenfelt. Active income earners 1943–1966 SOS original data. SCB employment recomputed 1 hour worked.

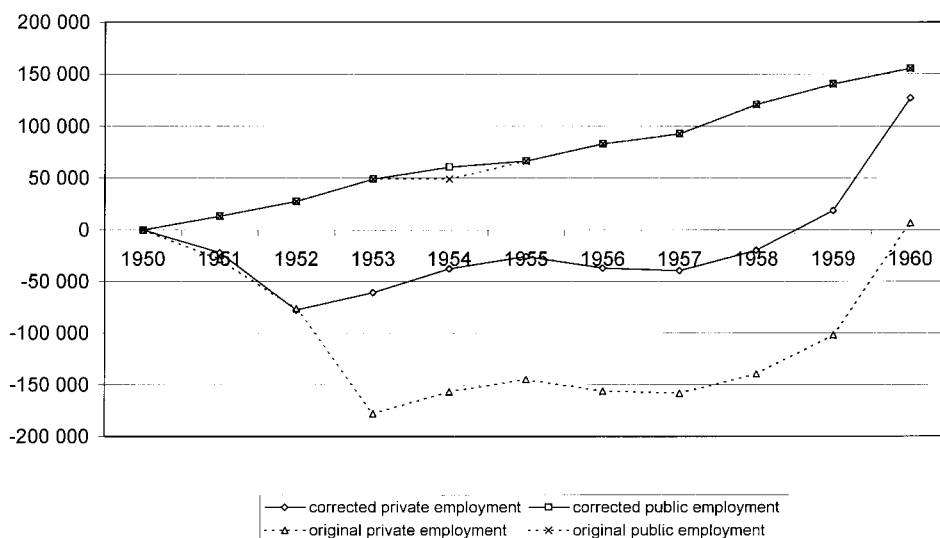


Fig. 3. Cumulated SCB employment change 1950–1960

Rather than an estimated near-zero increase in private employment 1950–1960 in the original extrapolation, the corrected series imply a more than four per cent increase. The original series leads to the faulty conclusion that nearly all employment growth took place in the public sector during the 1950s. Since this conclusion is consistent with the difference between the 1950 and 1960 Censuses it has been widely accepted among researchers. The around 125,000 individuals increase in private employment and the around 155,000 individuals increase in public employment are roughly comparable numbers, especially since the latter number is probably an overstatement.

The extrapolation arrives at 494,000 public employees in 1960, compared to 461,600 according to the National Accounts. The discrepancy is concentrated in SNI 9, where about 95 per cent of all public employees are reported. Average work-time in this industry primarily reflects public employment. Dividing worked hours by annual hours we exaggerate the number of public employees, indicating that the estimate of annual hours for public employees in SNI 9 is too low.

Total employment in 1960 according to the extrapolation is 3,561,384, and according to the National Accounts 3,615,700. The difference in level is partly due to different coverage between the wage-sum based older data and the survey based modern data, which include unpaid family helpers.

2.2.2. Bridging 1950 and 1951 in the SOS series

The SOS series present an entirely different set of problems. The basic raw data are fairly reliable, collected from first-hand sources, but the lack of overlap for the crucial year 1950 is troublesome. Not only is an overlap lacking but the information provided is actually a bit misleading.

For incomes below the 30,000 SEK limit, data from 1951 were collected only for those born on the 5th, the 15th, and the 25th of the month. The industry classification and income of co-taxed spouses not born on these dates were also noted. The survey population was

multiplied by 10 to arrive at the population totals, but the year is 365 days rather than the 360 days assumed by using the factor 10.

Moreover, the source notes mention that the sample of individuals born on the 5th and the 25th are relatively under-counted by around 1.5–2 per cent. That is, compared to those born on the 15th (given the reasonable assumption that births should be uniformly distributed over these dates). The reason given is that for those born on the 15th of the month names were also collected and used for other purposes. Therefore the local collectors presumably considered these data more important. Taking that, too, into account, the multiplying factor should be around 10.25.

Compared to earlier data the 1951 numbers underestimate the population of active income earners by more than 80,000. This problem is relatively simple to solve once it is detected. But it would not be obvious from the data since the discontinuity between 1950 and 1951 is a jump upwards not downwards as this correction implies.

The upward jump reflects the most problematic change in definitions. Co-taxed spouses were counted as one income earner up to 1950, but are counted separately in the sample survey, contingent on belonging to the sample population.¹¹ The 1952 tax assessment (of income earned 1951) states that 476,139 co-taxed spouses were added to the total population. Deducting this from the 1951 number gives 3,339,082 income earners by the old definition, which is 62,311 less than in 1950. This is hardly believable for a year that by no other account is a recession year.

The number of co-taxed spouses was also underestimated. In the source text it is assumed that each spouse has a ten per cent probability of being sampled. Then 20 per cent will be in the survey sample, from which must be deducted the one per cent of cases where both belong to the sample, which is to say that 100/19 or 5.263 times the co-taxed in the sample gives the total number of the co-taxed individuals in the population. Using a more correct probability for a co-taxed couple to be recorded in the sample would yield about 5.388 as a multiplier. Correcting for both underestimates results in a more reasonable increase between 1950 and 1951 of 21,777 income earners by the old definition.

The major problem is to distribute the 1951 addition of the co-taxed over industries and occupational status in order to adjust the 1951 industry data to the 1950 definition. There is no direct information on this distribution in the source. In SOS 1952 there is plenty of information on the characteristics of the co-taxed divided into males and females in a number of text tables. At first glance this might seem to be just the information needed. But a closer reading reveals that this information, unfortunately, refers to the distribution of all individuals that were co-taxed. The individuals that actually were added to the survey sample are only a sub-sample of this distribution.

This sub-sample is not likely to be distributed in the same way since it is the lower-income spouse – in general the female – that is added, and a large number of those have very low incomes indicative of low intensity work, seasonal or part-time. Experiments using these distributions to allocate the added individuals confirmed that suspicion by yielding quite unreasonable results.

¹¹ This means that the tax assessment records are included in the sample if either of the co-taxed spouses is born on the 5th, the 15th, or the 25th. In contrast with what is the case regarding the total survey this results in the inclusion of spouses with an income lower than the taxable income limit. The sample is, however, drawn from the same population of tax records, so no corrections need be made for that.

Without starting from scratch to construct new estimates from primary sources, only a judgmental approach was available to distribute the added spouse by industry and occupational status. By subtracting the 1950 active income earners from the 1951 numbers we could estimate a lower limit on the number of co-taxed inactive income earners. Guessing that an overwhelming majority of the added inactive income earners were females, we could make a reasonable guess as to how many of the added females were inactive. A small number of inactive males – that would not have determined the occupational status – is offset by the also small number of inactive females – that would have determined the occupational status – so the error should be relatively small.

This leaves us with an estimate of an addition of 398,431 active co-taxed individuals. The increase in active income earners is then 12,585, which is a little less than the KI estimate of an increase in total employment of 19,000 in 1951.¹² The estimate is rather conservative, since the guess about added inactive income earners may well be a bit too high. But the number is close enough to the contemporary source not to warrant any further adjustments.

The distribution over industries was then made in two steps. First, large increases in a category between 1950 and 1951 were taken to indicate large additions of co-taxed individuals. Even numbers were then deducted from this category. The remainder of the estimated 1951 addition of active co-taxed income earners was then distributed over industries and subtracted in line with the general trends that our other sources indicated. Occupational status and industry distribution were adjusted to reflect general information in SOS that low-income co-taxed females would include few self-employed and be more concentrated in manufacturing, trade and public administration.

One remaining but fairly simple problem was that the construction industry was included in manufacturing before 1952. We simply extrapolated the 1952 shares backwards, since the information available to compute any changes in the distribution was too unreliable.

2.3. *Linking classifications*

We used the SNI 69/77 code as a standard to link consistently to the National Accounts (see Table 2 for an overview of the differences). The 1950 and 1960 Censuses provide the only direct opportunity to recompile older definitions and were used to provide keys to earlier industry classifications. We only try to provide a rough basis for translating Jungenfelt's industry classification into SNI codes in 1950 using the 1950 Census and for the translation of SOS industries into SNI codes 1960, using the 1960 Census. Below we first describe the available information and then proceed to discuss how it was used for recompilation and adaptation to other classifications.

2.3.1. Census information

We chose to ignore occupational status in the recompilation, since definitions differ considerably between sources and cannot be compared to the census definitions. That

¹² Reports from the National Institute of Economic Research (Meddelanden från Konjunkturinstitutet) series B: 14, p. 19.

subdivision should ideally be linked directly between the series rather than being benchmarked on the censuses.

There are several problems to consider. First, there is a general lack of comparability between the two censuses, which in part has to do with differences in the methods used to establish the industry classification for the individuals.¹³ This we cannot do much about. The other part is, however, changes in industry classification where the 1960 Census was adapted to ISIC classifications.¹⁴ The industry classification used in the 1950 Census is close to the industry definitions used by Jungenfelt and in SOS. The main problem for us was, thus, the links to the SNI 69/77 classification, which differs considerably from the older variants.

In 1960 SCB performed specially designed control surveys to establish how the AKU employment relates to the census gainful employment.¹⁵ The main differences are part-time work (not in the census); absentees (in the census but not in the AKU); drafted military personnel (not in the census but in the National Accounts). The control survey also found a substantial under-count of about 200,000 individuals – mainly family helpers in agriculture – that should have been in the AKU.

2.3.2. Recompilation of the censuses

The recomplings mostly concern the 1950 Census, except for wholesale and retail trade in the 1960 Census and the division between public and private employment.¹⁶ The 1960 Census provides a (partial) recompilation of the 1950 Census. We use the 1950 Census to recompile service industries into major SNI 69/77 industries. Public employment was recompiled using definitions close to Jungenfelt's public administration. Our main problem regarding 1960 was to redistribute commercial and other services and separate public services from private services.

Banks, insurance, etc., real estate, and business services were reallocated from the commerce sector to a separate major industry corresponding to SNI 8. SNI 9, other personal services, includes paid domestic work – a separate industry earlier – but also substantial parts of what was previously called public services (services to the public, rather than government-provided services).

Compared to the original 1950 Census classification, SNI 1, 4, 5, 7 and 8 can be considered more or less equivalent to original industries or major components of these. SNI 2, mining and quarrying, includes components relating to stone and earth industries that cannot be identified in the earlier classification, so a ratio distribution link was applied for the separation of SNI 2 and SNI 3. SNI 6 also needs to be linked by census ratios since some minor components differ even after deduction of the SNI 8 industries. That is also the case for public services which includes several private components in the original definition that cannot be directly deducted.

¹³ In the 1960 Census, Part XI, Appendix 1, p. 11 f, these differences are discussed in detail.

¹⁴ The 1960 Census Part VI, Appendix 2 contains a conversion table between the 1960 and the 1950 Census classification.

¹⁵ A short summary of the results can be found in Statistical Reports N 1975:98, and more detail in Statistical Reports N 1970:44.

¹⁶ The recomplings are based on the 1950 Census, Part VI, Table 1, pp. 81–142, table G, p. 22; and the 1960 Census Part IX, Table 3, pp. 10–21 and Part VI, Table A, p. 4.

Table 2. Specification of differences between industries in different SNI codes

Classification of industries according to:

SNI 60	SNI 69	SNI 77
0. Agriculture, hunting, forestry and fishing	1. Agriculture, hunting, forestry and fishing	1. Agriculture, hunting, forestry and fishing
1. Mining and quarrying	2. Mining and quarrying	2. Mining and quarrying
2–3. Manufacturing food, beverages, tobacco textile, wearing etc wood, furniture etc paper and pulp etc printing and publishing etc leather industries nonmetallic minerals chemical industries etc metal industries other industries	3. Manufacturing food, beverages, tobacco textile, wearing etc wood, furniture etc paper and pulp etc printing and publishing etc leather industries nonmetallic minerals chemical industries etc metal industries other industries	3. Manufacturing food, beverages, tobacco textile, wearing etc wood, furniture etc paper and pulp etc printing and publishing etc leather industries nonmetallic minerals chemical industries etc metal industries other industries
4. Construction	4. Electricity, gas and water	4. Electricity, gas and water
5. Electricity, gas and water	5. Construction	5. Construction

6. Wholesale and retail trade, banks,
insurance, and real estate
wholesale and retail trade
banks, insurance etc
real estate

7. Transport, storage and communication

8. Services
public administration
education services
health services
social services
business and private personal services,
restaurants and hotels etc

6. Wholesale and retail trade, restaurants
and hotels
wholesale and retail trade
restaurants and hotels

7. Transport, storage and communication

8. Financing, insurance, real estate and
business services
banks, insurance etc
real estate
business services

9. Public and personal services

public administration
education services
health services
social services
other personal services etc

6. Wholesale and retail trade, r
and hotels
wholesale and retail trade
restaurants and hotels

7. Transport, storage and comm

8. Financing, insurance, real es
business services
banks, insurance etc
real estate
business services

9. Other private services

A. Private business
B. Public services
administration
health services
education services
social services
religious services

The main purpose of recompiling the 1950 Census is to provide links to Jungenfelt. Manufacturing and handicraft in Jungenfelt includes all of SNI 2, 3 and 4. Linking to SNI industries in 1950 requires distributing these industries according to the ratios in the 1950 Census. Commerce and other private services in Jungenfelt includes SNI 6 as well as some components of SNI 8 and 9, but only banks and insurance in SNI 8 are separately accounted in Jungenfelt (1959). Real estate and business services are included in professional services with other personal services. Public administration in Jungenfelt is, however, equivalent to the SNI definition of public services.

The 1960 Census was recompiled into SNI 69/77 industries to provide links to the SOS industry classification. In the SOS data no finer subdivisions are available, so we cannot determine the correspondence between industry classifications as well as we can for Jungenfelt (1959). The classification described in the source notes¹⁷ is short and somewhat vague. As far as we can determine, SOS manufacturing includes SNI 2, 3 and 4. SOS commerce only needs to be distributed between SNI 6 and 8, using ratios given by the 1960 Census. SOS professionals and household services make up the bulk of private SNI 9, but also a part of SOS public services, e.g., private health services, belongs here. We cannot exclude this so the most straightforward solution is to aggregate professionals and household services and ratio link to SNI 9.

For the reallocations into SNI 2, 3, and 4 we mainly used the recompilation in the 1960 Census. Mining and quarrying activities are, however, included in two different sub-industries in the 1950 Census. We reallocate an average of mining and quarrying from SNI 3 to SNI 2, and the remaining discrepancy (around 10,000 individuals) to the category other industries in SNI 3.

We also had to reallocate some components between SNI 6, 8 and 9: pharmacies from other personal services to SNI 6; a discrepancy in laundries and cleaning from trade to SNI 9; banks, insurance, real estate and business services from trade to SNI 8. We also exclude public services in SNI 69/77 from the census definitions. In principle, we have taken the sub-industries used by Jungenfelt as components of public administration, and from the detailed census tables deducted all categories that are likely to be mainly private services. From all other categories we deducted any persons classified as self-employed or family helpers, since no such persons should be included in public employment. The result is very close to public administration in Jungenfelt.

3. How Reasonable Are the Constructions?

In this section we assess the validity of the constructions in Section 2 by making comparisons. Some differences that we detect are due to different definitions that we cannot reconcile. Other differences derive from data construction and differences in coverage between sources. We point to possible remedies but abstain from making corrections.

3.1. Comparisons for validation of linking

We make three different comparisons. First the SCB series is compared to the SOS series reclassified to SNI 69/77. This provides a validation of the industry links in 1960. Second,

¹⁷ E.g., in SOS 1951, p. 38* f.

we compare the recomputed SOS series with original classifications with the Jungenfelt (1959) series in the overlap between the two. This mainly provides a validation of the bridging of the 1950–1951 gap in the SOS series. Third, we compare both the SCB series and the SOS series with contemporary KI estimates in the 1950s. This provides a validation of how changes in the series rather than levels correspond to perceived business cycle variation in the 1950s. We cannot validate the 1950 industry links beyond what we have already done in the recompilation of the census material.

3.1.1. Comparing SCB with SOS industries

In Figure 4 we compare the SOS series – 1950 adjusted to individual active income earners – with the SCB series using the SNI classification and the industry links in 1960. We have here chosen to ratio link both series to the 1960 Census gainful employment to facilitate comparisons.

In SNI 1, agriculture, the difference is a slower downturn in SOS employment. Due to the large self-employment in this industry the extrapolation method in the SCB series should be more unreliable but total change 1950–1960 only differs by a few thousand persons. In SNI 2, 3, and 4 it may be a bit too crude to just extrapolate the 1960 ratios backwards in the SOS data. But again the differences are very small and the main difference is in the movements between 1950 and 1960.

In SNI 5, construction, differences are a little larger, but recall that construction in SOS is extrapolated backwards from 1952 and may overstate the number of active income earners in 1950. In SNI 6, trade, there is a faster increase in SOS that may reflect the income limit bias, but it could also be attributable to our distribution of commerce in SOS into SNI 6 and 8.

In SNI 7, transports, there is a major difference in evolution in the last few years. We have no well-founded explanation for this. Judging from the KI data it seems likely that the SCB series is the more accurate. The downturn in active income earners may be an artefact. No change in classification was reported, but it is nevertheless possible that subtle changes in collection and coding procedures may reallocate individuals to other industries. Public services in 1959 increase by around 22,000 persons, and some of the components in transport are public enterprises and authorities.

SNI 9, other services, show large discrepancies, and completely different trends. The SOS series is dominated by the strong decrease in household work. Private SNI 9 should include some parts of SOS public services but the appropriate ratio is difficult to determine.

Looking at the private business sector, the main difference is in the last two years, when the SCB series shows a much more rapid increase. Since the first AKUs were available in 1959, one might suspect that the last years of the hours-worked series may have been adjusted.

Public employment according to SOS shows the expected income limit bias except for the decrease 1951–1952. Commerce expands in spite of a general slowdown in the economy, so some public employees may have been reclassified there. The SOS definition of commerce does include some public activities like the central bank, pension boards and social insurance authorities.

The evolution of total employment is fairly similar, except for the increase in the SCB

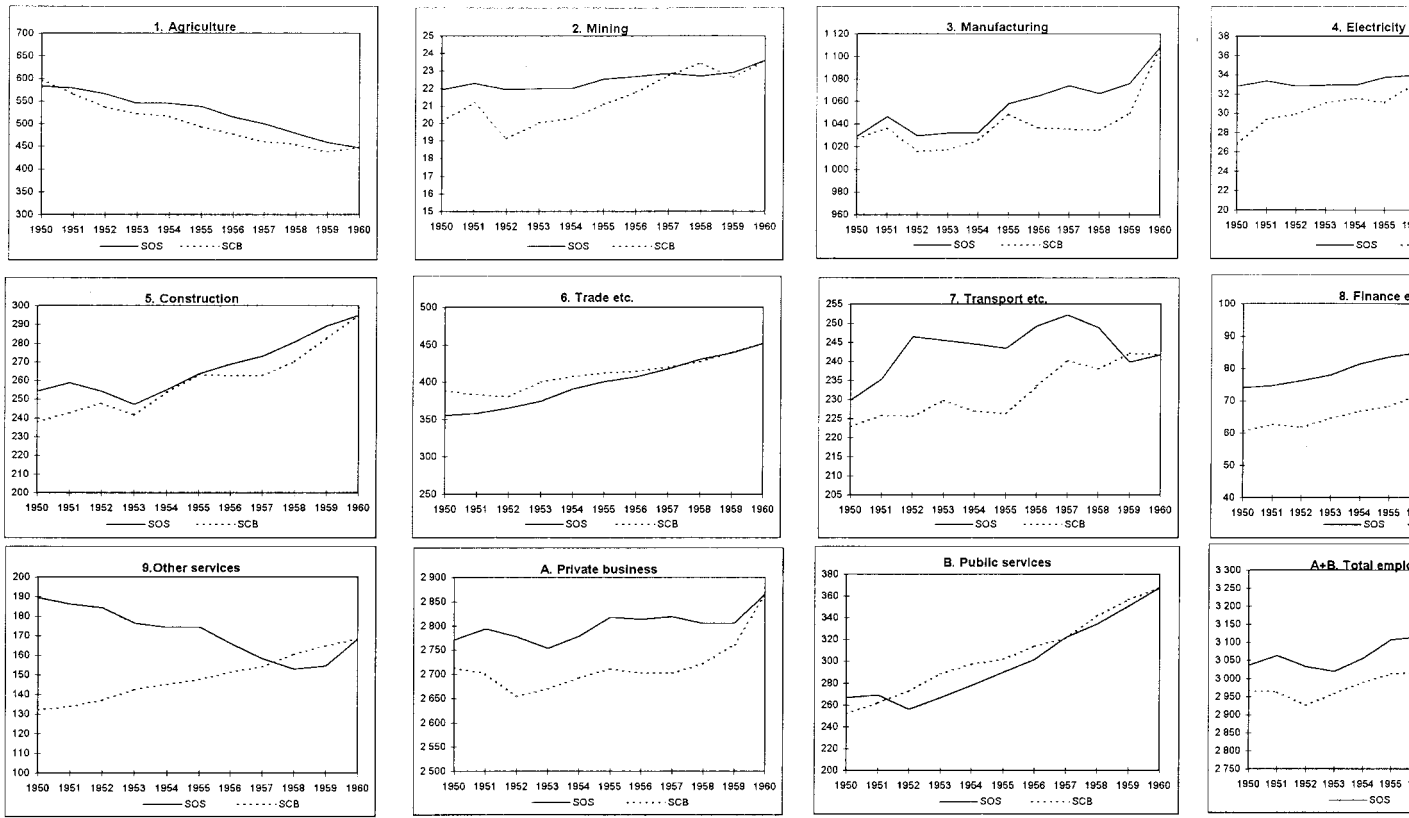


Fig. 4. SCB employment recomputed from hours worked and SNI industry-linked SOS active income earners. Ratio links to 1960 Census. 1,000's of ind

series the last few years caused by private business differences. It seems the SCB series are more reliable if all the SNI industries are needed. There is considerable structural change in the 1950s, so interpolation of the SOS ratio keys between the census benchmarks may improve on the SOS series in Figure 4. But we doubt that this would be sufficient to overturn the conclusion. For a coarser industry division (aggregating SNI 2, 3 and 4 as well as 6, 8 and 9) the SOS series is acceptable.

3.1.2. Linking SOS to Jungenfelt: a guiding comparison

The occupational status definitions in SOS and Jungenfelt are similar in most industries so we can compare occupational status in Figure 5.¹⁸ We also make use of the overlap 1943–1950, in some industries 1943–1955, to suggest improved links and adjustments to Jungenfelt's original series.

Jungenfelt reports a level of employment in 1950 that is some ten per cent too low to be comparable to later data. That discrepancy is not uniformly distributed over the industries. Jungenfelt's industry definitions are not identical to those of the taxation statistics but fairly close, so industry linking between these two sources is rather straightforward. Jungenfelt (1959) does not discuss the shift in method in the taxation statistics after 1950, although it is clear that co-taxed couples account for the bulk of the discrepancy in the total.

Jungenfelt includes private professional services in commerce, so we aggregate these two SOS industries. He also leaves out the nonspecified category. This is a nonnegligible quantity of between 20,000 to 50,000 individuals that we cannot distribute in any sensible way.

In agriculture the Jungenfelt employment data in the overlap are simply rounded numbers from the taxation statistics and only differ by the ratio link we have applied above. However, the period 1940–1943 in Jungenfelt has been linearly interpolated between figures derived from the 1940 Census and the 1943 active income earners (and then rounded). In manufacturing the estimates of employees show a growing discrepancy that may be due to the income limit effect in the taxation statistics. Jungenfelt's main source here is industrial statistics on the average number of employees in firms with more than five employees, complemented with less reliable data from surveys of smaller firms. Thus it is actually quite surprising that the differences are so small.

The self-employed differ considerably and have opposite trends in the overlap. One might suspect that our extrapolation of the 1952 ratio of manufacturing and construction is the cause. However, if we aggregate manufacturing and construction it does not help. The difference is thus more likely to be due to the definitions of self-employment. While SOS determines self-employment from income source, industrial statistics report only nonemployee managers as self-employed.

Although general trends for employees and self-employed are similar, the construction industry estimates illustrate another difficulty. The adjusted SOS series is much smoother, since Jungenfelt used average gross output productivity growth calculated from firm Censuses 1930 and 1950 to interpolate employment by dividing real gross output by

¹⁸ We only discuss the post-1940 part of the Jungenfelt series. Earlier parts of Jungenfelt differ substantially in construction.

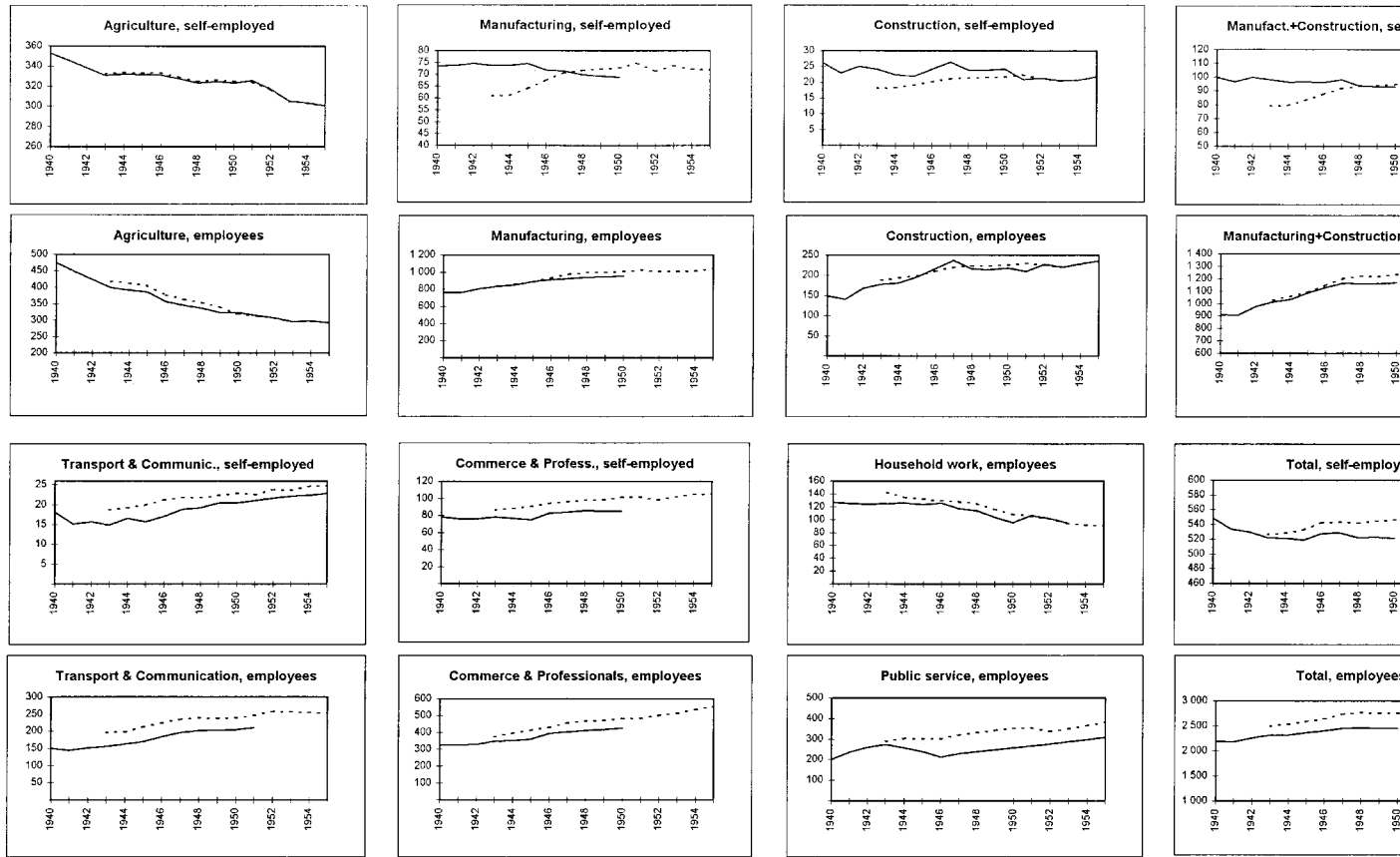


Fig. 5. Comparison between industry series from Jungenfelt (solid line) and adjusted SOS series (dotted line). 1,000 s of individuals

linearly interpolated productivity measures. Distribution over occupational status was inferred by the census distributions. This method yields excessively volatile employment series if labour productivity is procyclical. The transport and communications series in Jungenfelt have the same trend as SOS, but at a considerably lower level. The SOS definition is more comprehensive, including travel bureaus and tourism activities as well as sea pilots and lighthouse personnel, categories that are missing in Jungenfelt.

In commerce Jungenfelt includes private professionals, who are accounted separately in the SOS statistics. Jungenfelt has more employees but fewer self-employed than the original SOS series, reflecting perhaps that a considerable part of employees in commerce are married women and thus co-taxed. The adjusted SOS series does show a more similar distribution but at a higher total level. Jungenfelt's sources are liable to less than complete coverage and level errors.

Public employees differ considerably in the overlap. While Jungenfelt shows a sharp decline by some 60,000 individuals during the war years 1943–1946, SOS indicates a moderate increase. The around 100,000 people difference later on is almost entirely caused by the different trends during that period. Jungenfelt (1959, p. 66) states that employment has been directly calculated from the taxation statistics from the year 1943 but the numbers tell us otherwise.

Finally, the trend growth in total SOS employment is considerably higher than in Jungenfelt. The bias in active income earners from the taxation limit is one probable culprit, but part of the difference is due to the increasing difference between manufacturing employees in the sources and the puzzling war-year discrepancy in public employees. The SOS series seems an overall more reliable indicator of employment in most industries in the overlap, especially if the income limit bias is compensated. This suggests that ratio linking in 1943 could often be a satisfactory solution.

3.1.3. Check against contemporary estimates in the 1950s of employment changes

Finally we evaluated the SCB and SOS series by comparison with contemporary estimates of employment changes from KI.¹⁹ This was primarily intended to give indications as to which series mirrors business cycle variations better. The industry classification is similar to SOS. For a rough comparison with the extrapolated SCB series, SNI 2, 3, and 4 were aggregated to correspond to manufacturing, and SNI 6, 8 and 9 aggregated to correspond to commerce.

As we understand it – although we are unable to find detailed documentation – the KI figures have the status of rough and ready guesstimates based on statistics from unemployment agencies and union unemployment insurance funds. Mostly only rounded figures for the change in employment are given. The level of these numbers in general seems to be underestimated due to insufficient coverage. They can clearly not be taken too literally. The main purpose was to reflect the current business conditions so these employment figures can be expected to incorporate some such general information. This provides the check we need to evaluate with respect to business cycle variations.

¹⁹ KI published two series of periodical reports on the status of the Swedish economy starting in the early 1940s: Reports from the National Institute of Economic Research (Meddelanden från Konjunkturinstitutet), series A and B, later succeeded by The Swedish Economy (Konjunkturläget).

Comparing the general trend of changes in total employment the SCB series tend to lead the other series. As remarked above, this is due to the construction of the extrapolation. KI and SOS are similar in direction during the 1950s. The very strong employment increase in the last three years of the SCB series is not altogether trustworthy when compared with KI. The general picture of the business cycle seems rather well captured in the SOS statistics, contingent on KI having got it right.

The evidence becomes more mixed as we compare the separate industries, reminding us that similarities in aggregates need not be reflected in the components, where differences can be offsetting. These offsets may be coincidental but if they are repeated year after year it seems more likely that there is something in the series construction that drives the similarity in the aggregates. The KI and SOS changes are mostly more consistent with each other than the KI and SCB changes, but there are exceptions, for instance the anomalous decrease 1952 in SOS public services.

Only in construction do we find that both the other sources tend to move in the opposite direction to the KI series. The fact that construction work is highly seasonal and KI numbers are dependent on unemployment statistics may have something to do with it.

4. Concluding Remarks

We present and evaluate some of the existing information concerning the development of employment in Sweden during the 1950s in major industries. No published time series of employed individuals during the 1950s exist. We bridge a break in the taxation statistics on active income earners, correct unofficial hours-based extrapolations, and provide keys for linking older industry classifications to the SNI codes of official time series starting in 1960 or in 1950.

This work illustrates many common problems in the construction of historical time series. Taxation statistics provide us with a proxy that catches business cycle variation but within an obsolete industry classification. However, this classification is relatively simpler to link to the historical time series. An extrapolation to individuals on hourly employment data with a more modern industry classification makes industry links more reliable but at the cost of systematic bias in the variation of the series. Both approaches lead to similar estimates of the level of employment, thus validating each other in that respect. In both cases it was necessary to thoroughly check the original construction in order to clean the data of anomalies. The extrapolation was tainted by computational errors; the taxation statistics were compromised by a nonoverlapping break in method and definition. Original data in both cases leads to seriously misleading conclusions.

Although several obscure points remain, our work shows that total employment increased between 1950 and 1960 by around 250,000 people. Contrary to the impression from comparing the 1950 and the 1960 Census – where the around 150,000 people increase in gainfully employed is heavily concentrated in the public sector – the employment increase was more evenly divided between the public and private sector of the economy.

More generally our work illustrates many of the traps and trade-offs in linking historical

time series. Changes in definitions, classifications and data collection procedures cause serious problems, especially when there is no overlap between the old and the new series. Poor documentation may be misleading or make it impossible to understand the original construction of data. Computational errors can easily sneak into the material and change conclusions substantially.

But this case study also shows that it is possible to narrow down the uncertainty surrounding historical data quite considerably by comparing different sources armed with a thorough understanding of the material. Even more important, it shows that fairly rough extrapolations on existing information can be quite informative and yield good estimates not too far off from contemporary sources based on first-hand information.

Researchers using this material, however, need to be cautious about the details of the construction of the estimates. The exact use of the links provided must be guided by the purpose of the investigation. For example, taking the variation in the extrapolated employment series literally could easily lead to faulty causal inferences. As another example, the interaction of nominal taxation limits and inflation exaggerates the employment growth estimated from taxation statistics. Changes in the nominal limits also cause artificial slumps that researchers need to be aware of.

5. References

- 1950 Census (Folkräkningen 1950), (1955). Part IV–VI. Statistics Sweden (SCB), Stockholm. (In Swedish).
- 1960 Census (Folkräkningen 1960), (1963). Part VI, IX, and XI. Statistics Sweden (SCB), Stockholm. (In Swedish).
- Gunnarsson, G. and Lindh, T. (1997). Swedish Employment in the 1950s – Filling the Lacuna. Working Paper 1997:13, Department of Economics, Uppsala University.
- Jungenfelt, K.G. (1959). Lönernas andel av nationalinkomsten (The Wages' Share of National Income). Mimeo. Department of Economics, Uppsala University, Uppsala. (In Swedish).
- Jungenfelt, K.G. (1966). Löneandelen och den ekonomiska utvecklingen (The Wage Share and the Economic Development). The Industrial Institute for Economic and Social Research (Industriens utredningsinstitut), Stockholm. (In Swedish).
- Official Statistics of Sweden, Income and Property, Tax Assessments (SOS Inkomst och Förmögenhet. Skattetaxeringarna), (1944–1967). Statistics Sweden (SCB), Stockholm. (In Swedish).
- Reports from the National Institute for Economic Research (Meddelanden från Konjunkturinstitutet), (1949–1957), Series B, National Institute for Economic Research (Konjunkturinstitutet), Stockholm. (In Swedish).
- Reports from the National Institute for Economic Research (Meddelanden från Konjunkturinstitutet), (1948, 1959), Series A, National Institute for Economic Research, Stockholm, 1948 and 1959. (In Swedish).
- Reports on Statistical Co-Ordination (Meddelanden i samordningsfrågor), (1969). Standard för svensk näringsgrensindelning (SNI) (Standard for Swedish Classification of Economic Activities), Statistics Sweden, Stockholm. (In Swedish).

- Silenstam, P. (1970). Arbetskraftsutbudets utveckling i Sverige 1870–1965 (The Development of Labour Supply in Sweden 1870–1965). The Industrial Institute for Economic and Social Research (Industriens utredningsinstitut), Stockholm. (In Swedish).
Statistical Reports, National Accounts (Statistiska Meddelanden, Nationalräkenskaperna), (1970). N 1970:44. Statistics Sweden (SCB), Stockholm. (In Swedish).
- Statistical Reports, National Accounts (Statistiska Meddelanden, Nationalräkenskaperna), (1975). N 1975:98, Appendix 5. Statistics Sweden (SCB), Stockholm. (In Swedish).
- The Swedish Economy (Konjunkturläget), (1961). National Institute for Economic Research, Stockholm, November 1960 and November. (In Swedish).

Received February 1998

Revised September 2000